KỶ YẾU HỘI NGHỊ KHOA HỌC CÔNG NGHỆ
QUỐC GIA LẦN THỨ VII

ISBN: 978-604-913-300-8

FAIR

NGHIÊN CỨU CƠ BẢN VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN

THÁI NGUYÊN, 19 - 20/6/2014

Proceedings of the 7th National Conference on Fundamental and Applied Information Technology Research (FAIR'7)

NHÀ XUẤT BẢN KHOA HỌC TỰ NHIÊN VÀ CÔNG NGHỆ

Improving localization precision of visual SLAM using Kalman filter

Quoc- Hung Nguyen[1], Hai Vu[1], Thi Thanh - Hai Tran[1], Quang - Hoan Nguyen[2]

[1] Research Institute MICA, HUST - CNRS/UMI 2954 - Grenoble INP- Hanoi University of Science and Technology
[2] Hung Yen University of Technology and Education

{quoc-hung.nguyen,hai.vu, thanh-hai.tran}@mica.edu.vn, quanghoanptit@yahoo.com.vn

ABSTRACT*— This paper describes a Visual SLAM (Simultaneous Localization And Mapping) system developed on an intelligent system. The proposed system aims to support navigation services for visually impaired people in indoor environments. Toward this end, we utilize the Fast Appearance-Based Mapping (FAB-MAP) algorithm that is an appearance-based place recognition method. Although FABMAP algorithm is reliable in the outdoor scenarios, it still needs further improvements in indoor environments where contain repetitive structure scenes and sensory ambiguity. Therefore, two improvements are proposed. Firstly, we propose a scheme to learn discriminative scenes from experimental environments. This is to build a robust visual dictionary associating FAB-MAP algorithm. Secondly, we utilize a Kalman Filter to update position of the vehicle (like a mobile robot). The Kalman Filter keeps track of an estimate of the uncertainty in the robots position and also the uncertainty in the recognized scene that has seen in the environments. By this way, a feasible navigation on a mobile robot is up and run. We do not mean this is a perfect solution, what we mean is that proposed system could serve more reliable navigation service to blind/visually impaired people in indoor environments.*

Keywords *— Visual Odometry, Place Recognition, FAB-MAP algorithms, Kalman Filter.*

I. INTRODUCTION

Autonomous localization and navigation are extreme desirable services of peoples who suffer from visual impairment problems. Most of commercial solutions are based on the Global Positioning System (GPS), WIFI, LIDAR, Ultrasound, or fusion of them. iNavBelt uses ultrasonic sensors to procedure a 120-degree wide view ahead of the user [19]. GuideCane has an ultrasonic sensor head mounted on a long handle [3] . The EyeRing developed by MIT's Media Lab., is a finger-won device that translates images into aural signals. Although such kind of devices are useful to blind/visually impaired people in some environments. The major drawbacks are that they only give limited kind of information, and required well-focused user control. Recent techniques in the computer visions and robotics community offer substantial advantages to overcome those limitations. This paper towards to these techniques by using visual sensors mounted on an intelligent system (like a mobile robot). The proposed system aims to solve two problems: 1. Understanding the current environments. 2. Self-localization of robot. Regarding the problem 1, a question is that "what does the world look like?". This question involves in the building map of the environments and robot's trajectory. In contrast to this, self-localization service relates to estimating a pose to a relative position on the created map. It is to answer the second question "Where am I?".

A visual SLAM replying on the visual appearance of distinct scenes is responsible for finding optimal solutions for both above problems: building, maintaining a map of to the robot's trajectory and estimating landmark positions. Recent approaches like FAB-MAP aim at reaching a high recall rate at 100% precisions. FAB-MAP [4] is a probabilistic appearance-based approach to place recognition. It builds on a visual vocabulary learned from SURF descriptors. A Chow Liu tree is used to approximate the probability distribution over these visual words and the correlations between them. This allows the system to robustly recognize known places despite visual ambiguity. FAB-Map 2.0 has been applied to a 1000 km dataset and achieved a recall of 3.1% at 100% precision (14.3% at 90 % precision respectively). Although FAB-MAP approaches are reliable recognition places in large-scale environments. For indoor environments, repetitive structure and sensory ambiguity constitute severe challenges for any place recognition system. In our real experiments in indoor environments, by setting threshold to reach a 100 % precisions, it is very difficult to obtain high recall rate (~ 14% at 100% precisions). Consequently, this leads to preventing detecting true positives. In this work, we argue a robust FAB-MAP that is reliable to recognize known places through autonomous operating in an intelligent system. We focus on two improvements. We first clearly define the visual dictionary of the scenes. As context of indoor environments, many scenes has repetitive structure, the visual dictionary needs including only representative scenes. Secondly, we deploy a Kalman filter to update current position of the vehicle (mobile robot). This function thanks to states of the vehicle (like velocity of the mobile robot, step-walks of the people).

For implementations, the proposed system has two phases. *The first phase* is an off-line process including two main functions: build the robot's trajectories and learning (indexing) places in the environment. We simultaneously collect visual data for the off-line process by a self-designed imaging acquisition system. For building the trajectories of the environment, we utilize a robust visual odometry proposed in [8]. This is interesting method because it is successful to build trajectory using only one consumer-grade camera. In order to learn places in the environment, we utilize so-called loop closure detections method [4], [14]. The main idea for learning the visited places is that loop constraints can be found by evaluating visual similarity between the current observation and past images where are captured in one (or several) trials. *The second phase* is an online process. An agent (such as vehicle, human) is required to mount/wear a mobile device camera. The current observation is matched to a place in the database which is learnt in the off-line phase. We then using a Kalman filter to update the current position of the vehicle.

We evaluate results of the improvements through travels of a mobile robot moving along corridors of a large building. The experimental results of the matching place on the created map are successful with 74% recall and 88% precisions. This results are beyond the performance of the original FAB-MAP for indoor environments. The Kalman filter help to update position of the mobile robot. Consequently, the guidance to the blind people through movement of the mobile robot is feasible. The paper is organized as follows: In Section I, we present motivations and outline our approaches. In Section II, we briefly survey related works. In Section III, we present our vision-based system for autonomous map building and localization. We report the experimental results on real data in Section IV. Finally, we conclude and give some ideas for future works.

II. RELATED WORKS

Developing localization and navigation assistance tools for visually impaired people have been received many intention in the autonomous robotics community [5] . Most of them involve in finding out efficient solutions to the positioning data that come from different sensory modalities such as GPS, laser, Radio Frequency Identification (RFID), vision or the fusion of several of them. Loomis et al. in [12] surveyed efficiency of GPS-based navigation systems supporting visually impaired people. The GPS-based systems share similar problems: low accuracy in urban-environments (localization accuracy is limited to approximately 20 m), signal loss due to multi-path effect or line-of-sight restrictions due to the presence of buildings or even foliage. Kulyukin et al. [10] proposed a system based on Radio Frequency Identification (RFID) for aiding the navigation of visually impaired people in indoor environments. The system requires the design of a dense network of location identifiers. Helal et al. [9] proposed a wireless pedestrian navigation system. They integrated several signals such as voiced, wireless networks, Geographic Information System (GIS) and GPS to provide the visually impaired people an optimized route.

Recent advanced techniques in computer vision offer substantial solutions with respect to localization and navigation services in known or unknown environments. The vision-based approaches are safe navigation and provide a very rich and valuable perception information of the environment. Alcantarilla [6] utilizes well-known techniques such as Simultaneous Localization and Mapping (SLAM) and Structure from Motion (SfM) to create 3-D Map of an indoor environment. He then utilizes means of visual descriptors (such as Gauge-Speeded up Robust Features, G-SURF) to mark local co-ordinate on the constructed 3-D map. Instead of building a prior 3-D map, Lui et al. [11] utilize a pre-captured reference sequence of the environment. Given a new query sequence, their system desires to find the corresponding set of indices in the reference video. Many specific applications that also are based on vision sensors are developed to support typical daily activities of the visually impaired people. For example, [2] develops an application, names *LocateIt*, which supports blind people locate objects in the indoor environments. In [22], *ShelfScanner* is a real-time grocery detection, that allows online detection of items on a shopping list.

Regarding to map building and localization services, visual SLAM has been proven to be quite successful in navigation for autonomous robotic systems [1]. By means of visual SLAM techniques, some wearable applications are proposed. Pradeep et al. [17] presents a head-mounted, stereo-vision for detecting obstacles in the path and warn subjects about their presence. They incorporate visual odometry and feature based metric-topological SLAM. Murali et al. in [13] estimate the users location relative to the crosswalks in the current traffic intersection. They develop a vision-based smart-phone system for providing guidance to blind and visually impaired travelers at traffic intersections. The system of Murali et al. in [13] requires supplemental images from Google Map services, therefore it is suitable with travels at outdoor environments only. With SLAM-based approaches, it is possible to build a map at the same time the location of the people who wears cameras standing/moving in the environment. However, the complexity of the map building task varies in function of environment size. In some case, a map can be acquired from visual sensor, but in other cases, the map is such that it must be constructed from other sensor modalities such as GPS, WIFI [4]. Furthermore, matching a current view to a position on the created map seems to be the hardest problem in many works [1], [7]. The appearance-based place recognition has been conducted by [20] who borrowed ideas from text retrieval systems and introduced the concept of the so called visual vocabulary. The idea was later extended to vocabulary trees by [15], allowing to efficiently use large vocabularies. [18] demonstrated city-scale place recognition using these tree structures. Recently, Maddern et al. report an improvement to the robustness of FAB-MAP by incorporating odometric information into the place recognition process. [21] propose BRIEF-Gist, a very simplistic appearance-based place recognition system based on the BRIEF descriptor, BRIEF-Gist is much more easy to implement and its performance is comparable with FAB-MAP.

In our point of view, an incremental map is able to support us improving matching results. Therefore, different from above systems, we create a rich map as good as possible through many travels. When new observations arrive, these new observations must be locally and globally consistent with the previous construction. These problems are able to solve through the loop closure algorithms [4], [14]. We pay much attentions in the creating visual dictionary procedures of the FAB-MAP algorithm. Utilizing the GIST features [16], a holistic representation of the natural scenes, the representative frames can be selected in order to construct a robust visual dictionary of the environments.

III. THE PROPOSED APPROACHES

A. *Imaging acquisitions system*

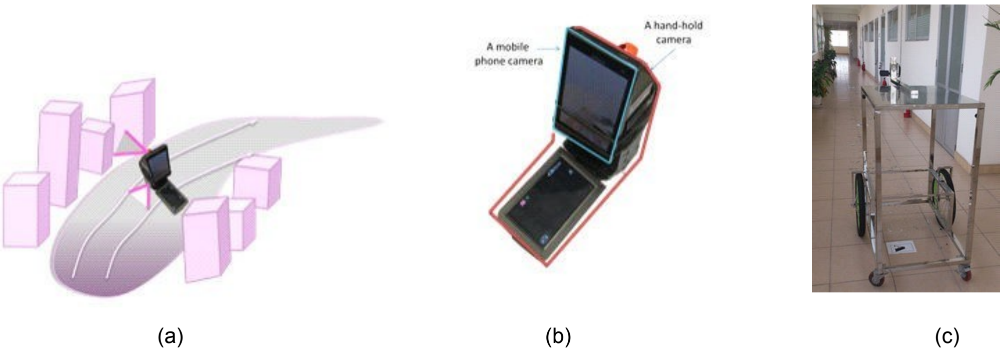

Fig. 1. (a) A schematic view of the visual data collection scheme. (b) The proposed imaging acquisition system in which a mobile phone camera is attached on rear of a hand-hold camera. (c). The image acquisition system attached on a wheel vehicle

We design a compact imaging acquisition system to capture simultaneously scenes and routes in the indoor environments. A schematic view of the data collection scheme is shown in Fig. 1(a). The proposed acquisition system has two cameras. One camera captures scenes around the environments. The second one aims at capturing road on the travels. The camera setting is shown in Fig. 1(b). These cameras are mount on a vehicle, as shown in Fig. 1(c). The collection data is described in Sec.IV.A

B. *The proposed framework*

General proposed system is shown in Fig. 2.

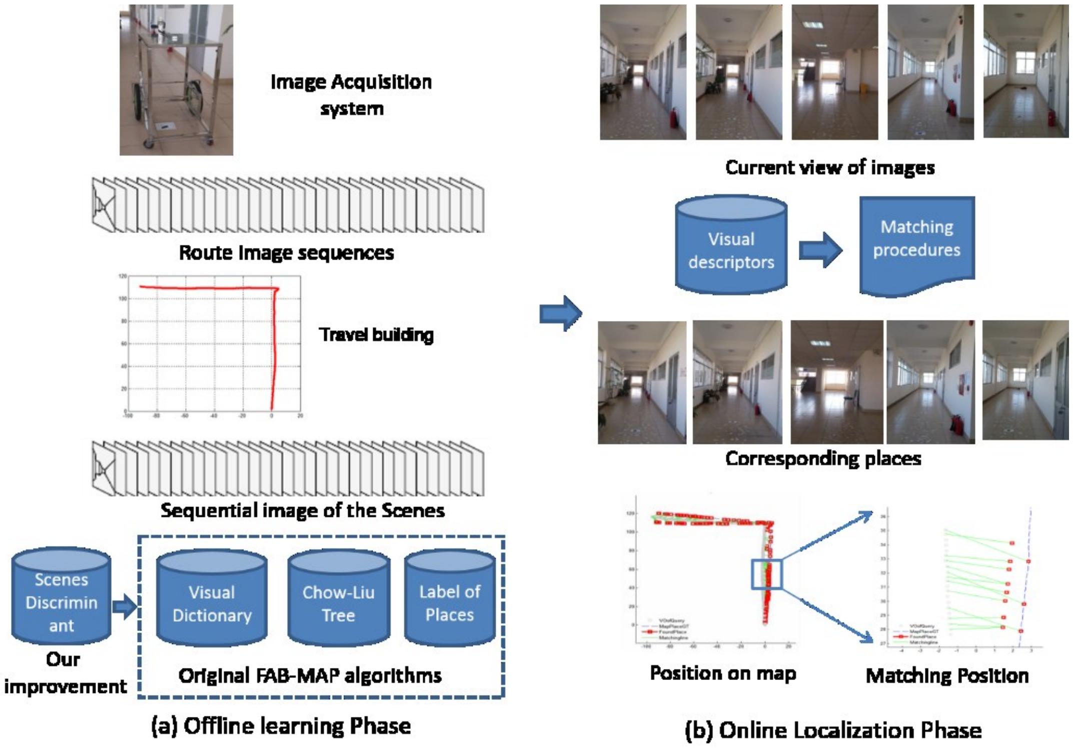

Fig. 2. The framework of the proposed system

The proposed system has two phases, as described below:

- **Off-line learning phase**: Using the collected visual data, this phase creates trajectories and learns the places along the travels. The techniques to construct the map and learning the places are described in Sec.III.C, respectively. Because scenes and route images are captured concurrently, the constructed map contains learnt places in corresponding positions of the travel.
- **Online localization**: A current view of image is described using a visual dictionary. These data associate matching the current view to a place what is labeled in the database through a probabilistic function. The current observation thus is able to match to a corresponding position on the constructed map.

C. *Matching image-to-map procedure*

The learning places from the sequential images that collected along trajectories aims at visually presenting appearances scenes. These visual presentations need to be easy implementation and efficient distinguishing scenes. To adapt with these issues, we involve the FAB-MAP technique [4] which is recently successful for matching places in routes over long period time. It is a probabilistic appearance-based approach to place recognition. Each time the image taken, its visual descriptors are detected and extracted. In our system, we utilize SURF extractors and descriptors for creating on a visual vocabulary dictionary. A Chow Liu tree is used to approximate the probability distribution over these visual words and the correlations between them. Fig. 3(a)-(b) shows the extracted features and visual words to build visual dictionary. Beyond the conventional place recognition approaches that simply compares image similarity between two visual descriptors. FAB-MAP involves co-occur visual word of same subject in the worlds. For example, Fig. 3(c) shows several windows subject, some of visual words are co-appearances.

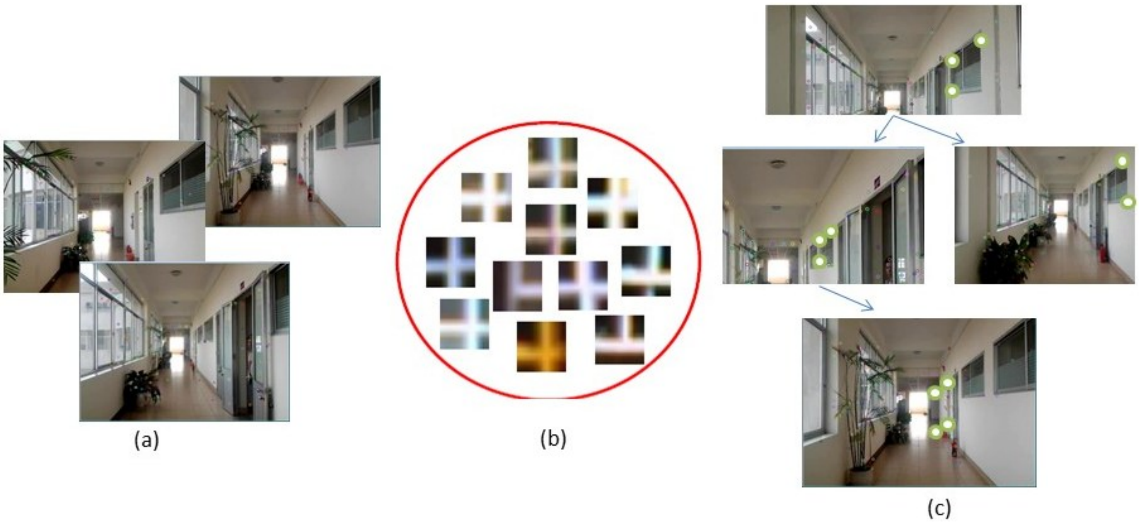

Fig. 3. FAB-MAP algorithm to learn places. (a) SURF features are extracted from image sequences. (b) Visual words defined from SURF extractors. (c). Co-occur of visual words by same object

Consequently, the distinct scenes are learnt from visual training data. For updating new places, we implement captured images through several trials.

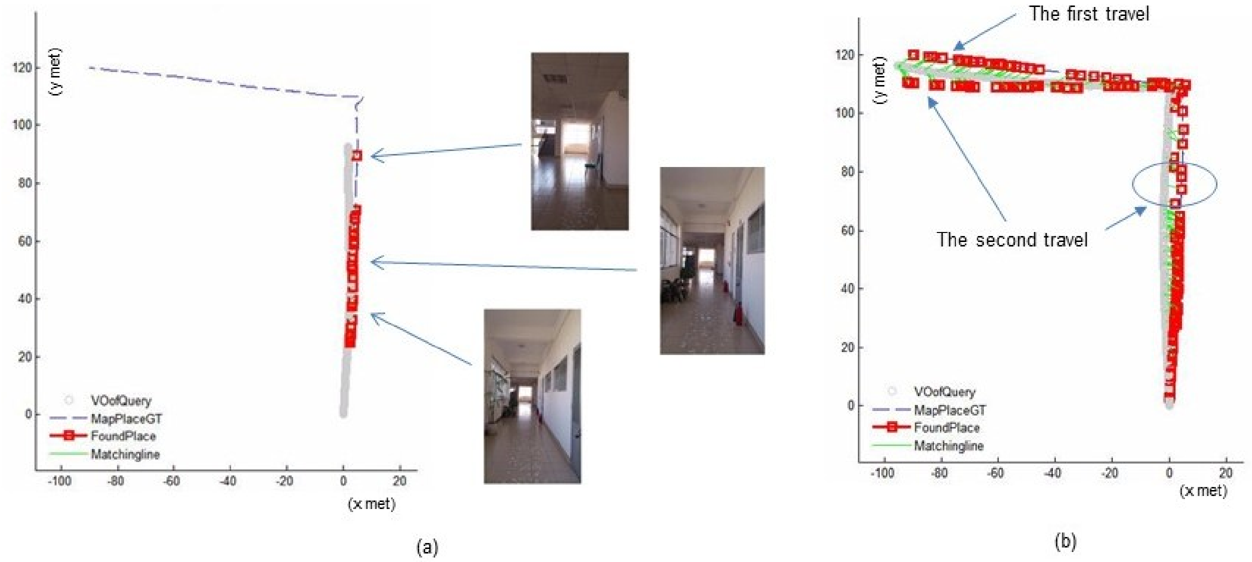

Fig. 4. (a) The places are learnt and their corresponding positions are shown in the constructed map data. (b) Many new places are updated after second trial

For each new trial, we compare the images with the previous visited places which are already indexed in a place database. This procedure calls a loop closure detection. These detections are essential for building an incremental map. Fig. 4 (a) shows only few places are marked by the first travel, whereas various places that are updated after the second travel as shown in the Fig. 4(b)

D. Distinguishing scenes for improving FAB-MAP's performances

Although related works [8], [6] report that FAB-MAP obtains reasonable results for place recognition over long travels in term of both precisions and recall measurements. However, those experiments were implemented in outdoor environments which usually contain discriminate scenes. Original FAB-MAP [2] is still unresolved problems of discriminating scenes to define visual dictionary. This issue affects to results of FAB-MAP when we deploy it in indoor environments, where scenes are repetitive structure and ambiguous. Therefore, a pre-processing step is proposed to handle these issues. A discriminative scenes based on holistic descriptors is deployed. The descriptors describe the appearance of the complete scene and not of single points in it. The idea of a holistic scene descriptor is not new and was e.g. examined by Oliva and Torralba [18] [17] with the introduction of the GIST descriptors. This global image descriptor is built from the responses of steerable filters at different orientations and scales.

Given a set of scene images S={I_1, I_2,….., I_n} we learn key frames from S by evaluating similarity of inter-frames. A feature vector F_i is extracted for each image I_i. In this work, the GIST feature [2] is utilized to build F_i. GIST presents a brief observation or a report at the first glance of a scene that summarizes the quintessential characteristics of an image. Feature vector F_i contains 512 responses which are extracted from an equivalent of model of GIST proposed in [11]. A Euclidean distance D_i between two consecutive frames is calculated to measure dissimilarity. Fig. 5(a) shows distance D_i of a sequence including 200 frames. The key-frame then is selected by comparing D_i with a pre-determined threshold value T. Examples of selecting two key-frames are shown in Fig. 5(b)

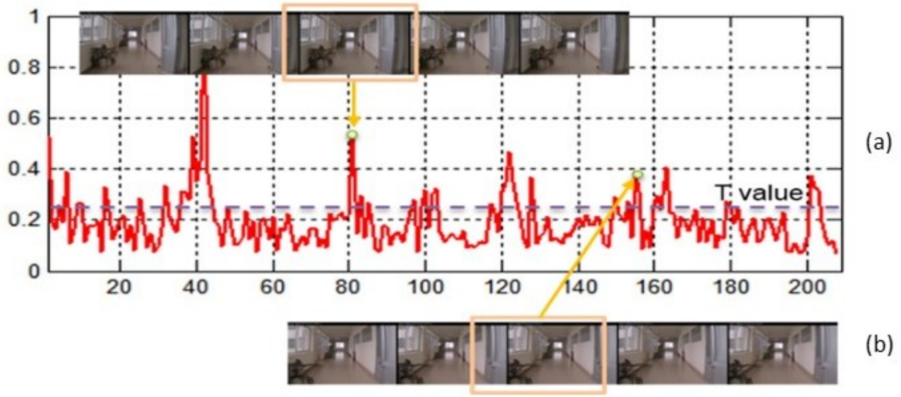

Fig. 5. (a) Dissimilarity between two consecutive frames. A threshold value T = 0.25 is pre-selected. (b) Two examples shows the selected key frames and their neighbor frames

E. Localizing a place to visited one in the constructed map

Given a current view, its position on the map is identified through a place recognition procedure. We evaluate the current observation at location L_i on the map by its probability when given all observations up to a location k:

$$\rho\left(L_i\middle|Z^k\right)=\frac{\rho(Z_k|L_i)\rho\left(L_i\middle|Z^{k-1}\right)}{\rho\left(Z_k\middle|Z^{k-1}\right)} \tag{1}$$

Where Z_k contains visual words appearing in all observations up to *k-1*; and Z^k presents visual words at current location *k*. These visual words are defined in the learning places phase. A probability $p(Z_k|L_i)$ infers observation likelihood that learnt in the training data. In our system, a L_i is matched at a place k∗ when $argmax(p(Z_k|L_i))$ is large enough (through a pre-determined threshold T = 0.9). The Fig. 6 shows an example of the matching procedure. Given an observation as shown in Fig. 6(a), the most matching place is found at *placeID* = 12. The probability $p(L_i|Z^k)$ is shown in Fig. 6(c) with a threshold *value = 0.9* whose the maximal probability is *placeID = 12*. A confusion matrix of the matching places for an image sequence is shown in Fig. 6(d). Although the confusion matrix shows that we can resolve almost places in a testing phase, some misrecognition places makes troubles for navigation services. To solve some, we introduce a Kalman filter to update current positions the vehicle, in which the current matching place is constrain in an uncertainty model. The velocity of the robot is known in this case.

F. The Kalman Filter (KF)

In our context, the observations of the robot are images captured over time, which then be converted to coordinates (x, y, z) in a predefined coordinate system using above matching procedure. However, in indoor environment, the scene does not change enough. Consecutive scenes could repeat when the robot moves. Therefore, the performance of image matching is not good. Sometimes, a current observation could be matched with a very far forward / backward image that makes incorrect localization of the robot.

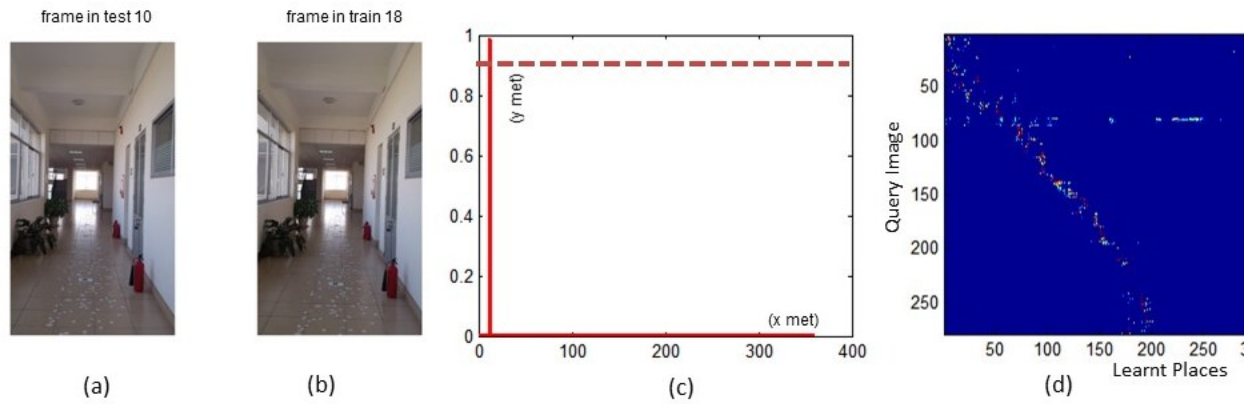

Fig. 6. (a) Given a current observation, (b) the most matching place. (c) The probability p(Li|Zk) calculated with each location k among K = 350 learnt places. (d) Confusion matrix of the matching places with a sequential collected images (290 frames).

To overcome this problem, we propose to use Kalman filter to correct the position of the robot from observation. Kalman filter is one of the most popular techniques to SLAM problem. It uses a series of observations over time to estimate unknown variables that are expected to be more precise than using single observations alone.

In our specific context, we give some notations as follows

- **State vector**: We suppose that the robot moves in a flat plane, so the z coordinate of the robot is constant then we can ignore it. The state of the robot at a given time k is simply presented by its coordinates and velocity in two directions x and y.

$$x = \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} \tag{2}$$

- **Observation vector**: At each time where the image matching is found, the position of the robot could be estimated. We use this information as observation in Kalman filter

$$z = \begin{bmatrix} x \\ y \end{bmatrix} \tag{3}$$

- State transition model F_k allows to predict the state vector at time *k+1* :

$$x_{k+1} = F_k * x_k + w_k \tag{4}$$

where w_k is process noise, which is assumed to follow a normal distribution with covariance Q_k: $w_k \sim N(0, Q_k)$. If the robot moves with constant vector, the simplest state transition model could be:

$$F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where Δt is the time duration of each iteration

- **Observation** model H_k maps the true state space into the observed space:

$$z_k = \mathrm{H}_k * x_k + \mathrm{v}_k \tag{6}$$

In our case

$$\mathrm{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where v_k is observation noise which is assumed to be zero mean Gaussian white noise with covariance R_k: $v_k \sim N(0, R_k)$

The Kalman filter works in two-step: prediction and update. At prediction step, Kalman filter predicts temporal evolution of the state. Once an observation incomes, these values will be updated using a weighted average.

At time t, the state of the filter is represented by two variables: a posteriori state estimate at time k given observations up to and including at time k $\hat{x}_{k|k}$ and a posteriori error covariance matrix (a measure of the estimated accuracy of the state estimate) P_k

- Prediction

$$\hat{x}_{k|k-1} = \mathrm{F_k} * \hat{x}_{k-1|k-1} \quad (7)$$

$$\mathrm{P}_{k|k} = \mathrm{F_k}\mathrm{P}_{k-1|k-1}\mathrm{F_k^T} + \mathrm{Q_k} \quad (8)$$

- Update

+ Innovation of observation: $$\tilde{y}_k = z_k - \mathrm{H}_k * \hat{x}_{k|k-1} \quad (9)$$

+ Innovation of covariance: $$\mathrm{S}_k = \mathrm{H_k}\mathrm{P}_{k|k-1}\mathrm{H_k^T} + \mathrm{R_k} \quad (10)$$

+ Optimal Kalman gain: $$\mathrm{K_k} = \mathrm{P_{k|k-1}}\mathrm{H_k^T}\mathrm{S_k^{-1}} \quad (11)$$

- Update state estimate:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \mathrm{K}_k\tilde{y}_k \quad (12)$$

- Update covariance estimate:

$$\mathrm{P}_{k|k} = (\mathrm{I} - \mathrm{K}_k\mathrm{H}_k)\mathrm{P}_{k|k-1} \quad (13)$$

IV. EXPERIMENTAL RESULTS

A. Evaluation Environments

- **Setting up environments:** We examine the proposed method in a corridor environment of a building, where is 10th floor of International Research Institute MICA-Hanoi University of Science and Technology (HUST).
- **Database collection:** Two camera devices are mount into a vehicle as shown in Fig. 1(c). A person moves at a speed of 1.25 foot/second along the corridor. The total length of the corridor is about 60 m. We collect data in four times (trials), as described in Table 1

Table 1. Three rounds data results

Trials	Total Scene images	Total road images	Duration
L1	8930	2978	5:14
L2	10376	2978	5:30
L3	6349	2176	3:25
L4	10734	2430	4:29

B. Experimental results

We evaluating the proposed system with aspects of the place recognition rate on the created map. To define visual word dictionary as described in Sec.III.C, we use collected images from L1 trial. About 1300 words are defined in our evaluation environments. We then use dataset from L4 travel to learn place along the travel. Totally, K = 140 places are learnt. The visual dictionary and descriptors of these places are stored in XML files. The collected images in L2 and L3 travels are utilized for the evaluations. Visually, some matching places results from L3 travel are shown in Fig. 7.

Two demonstrations are shown in details in Fig. 7 (around position A and position B). Case A shows a query image (from L3 travel) is matched to a learnt place. Therefore, its corresponding positions on the map is able to localize. A zoom-in version around position A is shown in the top panel. Case B show a *"no place found"* that query image was not found from learnt place database. For the qualitative measurement, we then evaluate the proposed system using two criteria: Precision is to measures total place detected from total query images, whereas Recall is to measure correct matching places from detected places. We setup a predetermined threshold for matching place (T = 0.9). Table 2 shows precision and recall with L2 and L3 travels with/without scene discriminant step. For learning place (using original FAB-MAP, without scene discrimination), the recall of L3 travel is clearly higher than L2. The main reason is that some "new" places where were not learnt from L4 are able to update after L2 running. Therefore, more "found" places is ensured with L3 travel.

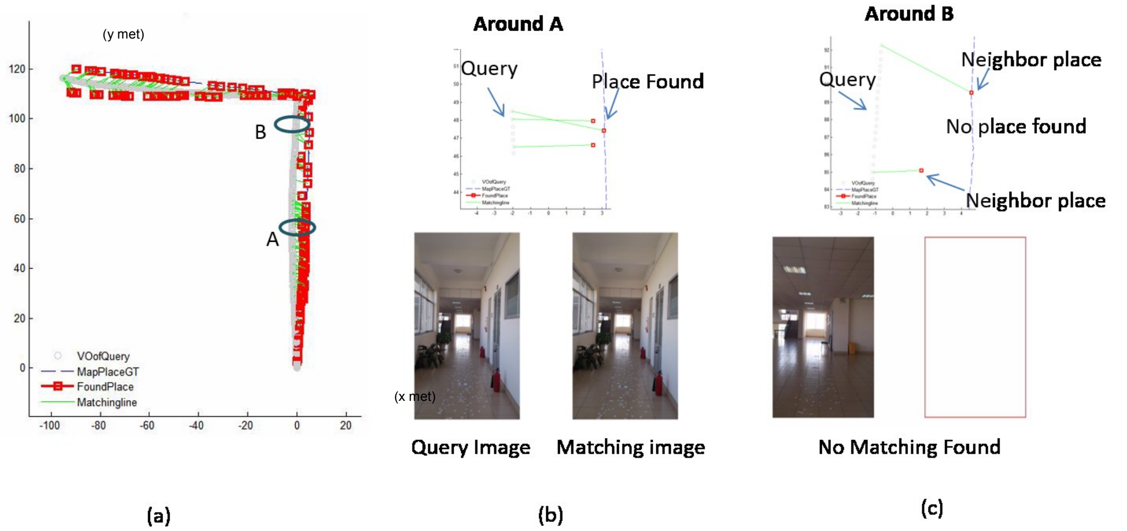

Fig. 7. (a) Results of the matching image-to-map with L3 trial. Two positions around A and B are given. (b)-(c): current view is on the left panel (query image); matching is on the right panel. Upper panel is a zoom-in around corresponding positions.

Table 2 also shows efficient of scene discriminations step (Sec.IV.B) the performances of image-to-map matching obviously increasing and stable for precisions measurement with scene discrimination step, whereas high confidence of the recalls is still consistent.

Table 2. Result of the matching places (FAB-MAP algorithms) without and with Scene discriminations

Travels	Without scene discrimination		With scene discrimination	
	Precision	Recall	Precision	Recall
L2	12%	90%	67%	82%
L3	36%	85%	**74%**	**88 %**

To show effectiveness of the applying Kalman filter, Fig. 8 demonstrates navigation data without and with using Kalman filter. Using only results place recognition (Fig. 8 – left panel), the directions supporting navigation services obviously uncontrolled. Some matching place (show in numbers) are mess and unordered in this case. Main reasons are some places are wrong matching (e.g., place ID = 11, shown in bottom panel). By using Kalman Filter, directions supporting navigation services is ordered. We can clearly observe the effectiveness on Fig. 8 – right panel.

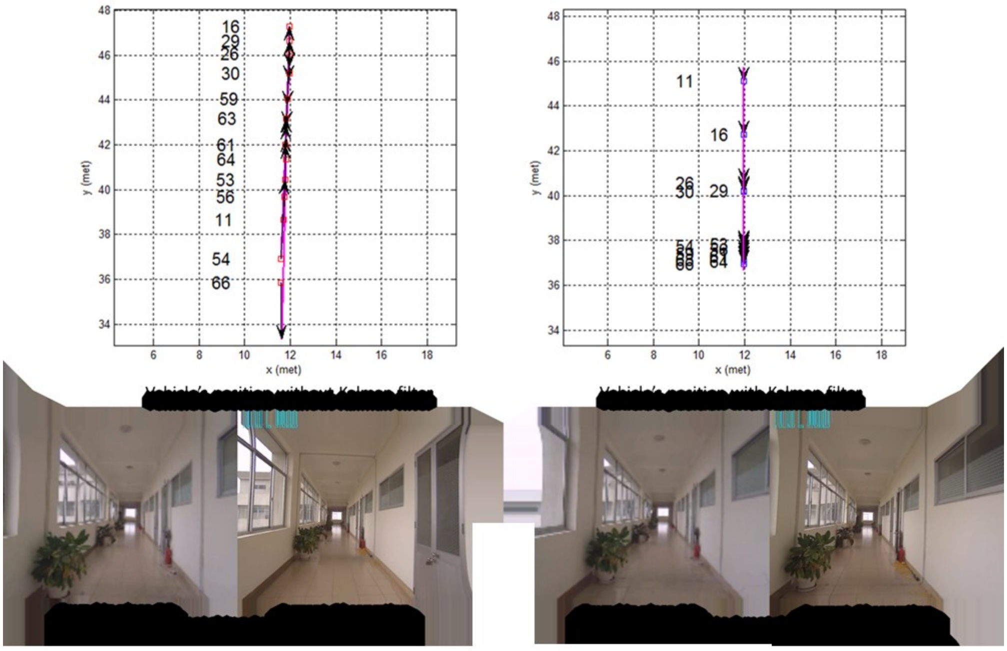

Fig. 8. Position of the vehicle without/with Kalman Filter. Top row: Left panel: some positions of the vehicle on the map using only results of the matching image-to-map procedures. The arrows show directions to guide vehicle. Numbers on left of each red box show placeID of the current observation. Right panel: poistions of the vehicle are updated using Kalman filter. Bottom row: Left panel: show wrong matching at placeID #11. This result yields wrong direction to vehicle. Righ panel: is a good matching at placeID = 56. No update with Kalman filter in this case.

V. CONCLUSIONS

In this paper, we presented a vision-based system for both autonomously map building and localizing services. We successfully created the map of the indoor environment using the visual odometry and learning places. We improved the FAB-MAP algorithms to solve indoor recognition problems. A visual dictionary is learnt using only representative scenes in the experimental environments. The results of matching image-to-map are high confidence for navigation service thanks to Kalman filter. The proposed system therefore is able to provide us deploying navigating services in the indoor environments. The proposed system directs to support blind/visually impaired peoples. The evaluations on the visually impaired/blind people direct us to future works.

REFERENCES

[1] Bailey T. and Durrant-Whyte H. (2006), "Simultaneous Localisation and Mapping (SLAM): Part II State of the Art,"

[2] Bigham J. P., Jayant C., Miller A., White B. and Yeh T., "VizWiz::LocateIt - enabling blind people to locate objects in their environment," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, 65-72.

[3] Borenstein J. and Ulrich I., "The guidecane-a computerized travel aid for the active guidance of blind pedestrians," in *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, 1997, 1283-1288.

[4] Cummins M. and Newman P. (2008), "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, 27, 647-665.

[5] Dakopoulos D. and Bourbakis N. G. (2010), "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40, 25-35.

[6] Fernández Alcantarilla P., "Vision based localization: from humanoid robots to visually impaired people," Electronics, University of Alcala, Ph.D. Thesis, 2011.

[7] Fraundorfer F. and Scaramuzza D. (2012), "Visual Odometry : Part II: Matching, Robustness, Optimization, and Applications," *Robotics & Automation Magazine, IEEE*, 19, 78-90.

[8] Hamme D. V., Veelaert P. and Philips W., "Robust visual odometry using uncertainty models," presented at the Proceedings of the 13th international conference on Advanced concepts for intelligent vision systems, Ghent, Belgium, 2011.

[9] Helal A., Moore S. E. and Ramachandran B., "Drishti: an integrated navigation system for visually impaired and disabled," in *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, 2001, 149-156.

[10] Kulyukin V., Gharpure C., Nicholson J. and Pavithran S., "RFID in robot-assisted indoor navigation for the visually impaired," in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, 2004, 1979-1984 vol.1972.

[11] Liu J. J., Phillips C. and Daniilidis K., "Video-based localization without 3D mapping for the visually impaired," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, 23-30.

[12] Loomis J. M., Golledge R. D. and Klatzky R. L. (2001), "GPS-based navigation systems for the visually impaired,"

[13] Murali V. N. and Coughlan J. M., "Smartphone-based crosswalk detection and localization for visually impaired pedestrians," in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, 1-7.

[14] Newman P. and Kin H., "SLAM-Loop Closing with Visually Salient Features," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2005, 635-642.

[15] Nister D. and Stewenius H., "Scalable recognition with a vocabulary tree," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, 2161-2168.

[16] Oliva A. and Torralba A. (2001), "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, 42, 145-175.

[17] Pradeep V., Medioni G. and Weiland J., "Robot vision for the visually impaired," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, 15-22.

[18] Schindler G., Brown M. and Szeliski R., "City-scale location recognition," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, 1-7.

[19] Shoval S., Borenstein J. and Koren Y. (1998), "Auditory guidance with the Navbelt-a computerized travel aid for the blind," *Trans. Sys. Man Cyber Part C*, 28, 459-467.

[20] Sivic J. and Zisserman A., "Video Google: A text retrieval approach to object matching in videos," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, 1470-1477.

[21] Sunderhauf N. and Protzel P., "Brief-gist-closing the loop by simple means," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, 1234-1241.

[22] Winlock T., Christiansen E. and Belongie S., "Toward real-time grocery detection for the visually impaired," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, 49-56.

Cải thiện độ chính xác định vị của hệ thống thị giác SLAM sử dụng bộ lọc Kalman

Nguyễn Quốc Hùng[1], Vũ Hải[1], Trần Thị Thanh Hải[1], Nguyễn Quan Hoan[2]

[1] Viện Nghiên cứu Quốc tế đa phương tiện MICA - Trường Đại học Bách Khoa Hà Nội
[2] Trường Đại học sư phạm Kỹ thuật Hưng Yên

{Quoc-Hung.NGUYEN, Hai.VU, Thanh-Hai.TRAN}@mica.edu.vn, quanghoanptit@yahoo.com.vn

Tóm tắt *— Bài viết này mô tả một hệ thống thị giác SLAM (đồng thời định vị và tự động xây dựng bản đồ) được phát triển trên một hệ thống thông minh . Hệ thống được đề xuất nhằm hỗ trợ các dịch vụ chuyển hướng để người khiếm thị trong môi trường trong nhà. Hướng tới mục tiêu này, chúng tôi sử dụng thuật toán Fast Appearance-Based Mapping (FABMAP) được dùng trong việc nhận dạng các vị trí xuất hiện trong một tập vị trí quan sát được. Mặc dù FABMAP là đáng tin cậy trong các tình huống ngoài trời, nó vẫn cần cải thiện hơn nữa trong các môi trường trong nhà, nơi nhận sạng vẫn còn là một vấn đề khó khăn. Vì vậy, có hai cải tiến được đề xuất. Thứ nhất, chúng tôi đề xuất xây dựng các phân đoạn khung cảnh đặc biệt, nổi trội khác biệt các môi trường thử nghiệm. Sau đó, xây dựng một từ điển trực quan mạnh mẽ bằng thuật toán FAB- MAP, toàn bộ hệ thống này được đặt lên robot. Thứ hai, chúng tôi sử dụng một bộ lọc Kalman để cập nhật vị trí hiện thời của robot. Với bộ lọc Kalman Filter theo dõi một ước tính không chắc chắn ở vị trí robot và cũng là không chắc chắn trong khung cảnh nhận dạng trong các môi trường thử nghiệm . Bằng cách này, một hướng khả thi về một robot di động tịnh tiến và đang hoạt động. Đây là một giải pháp mà chúng tôi đang trong thời gian tiến hành thực nghiệm, những đề xuất của chúng tôi có ý nghĩa phục vụ cho bài toán robot di động, chuyển hướng đáng tin cậy trợ giúp di chuyển của người mù và người khiếm thị.*